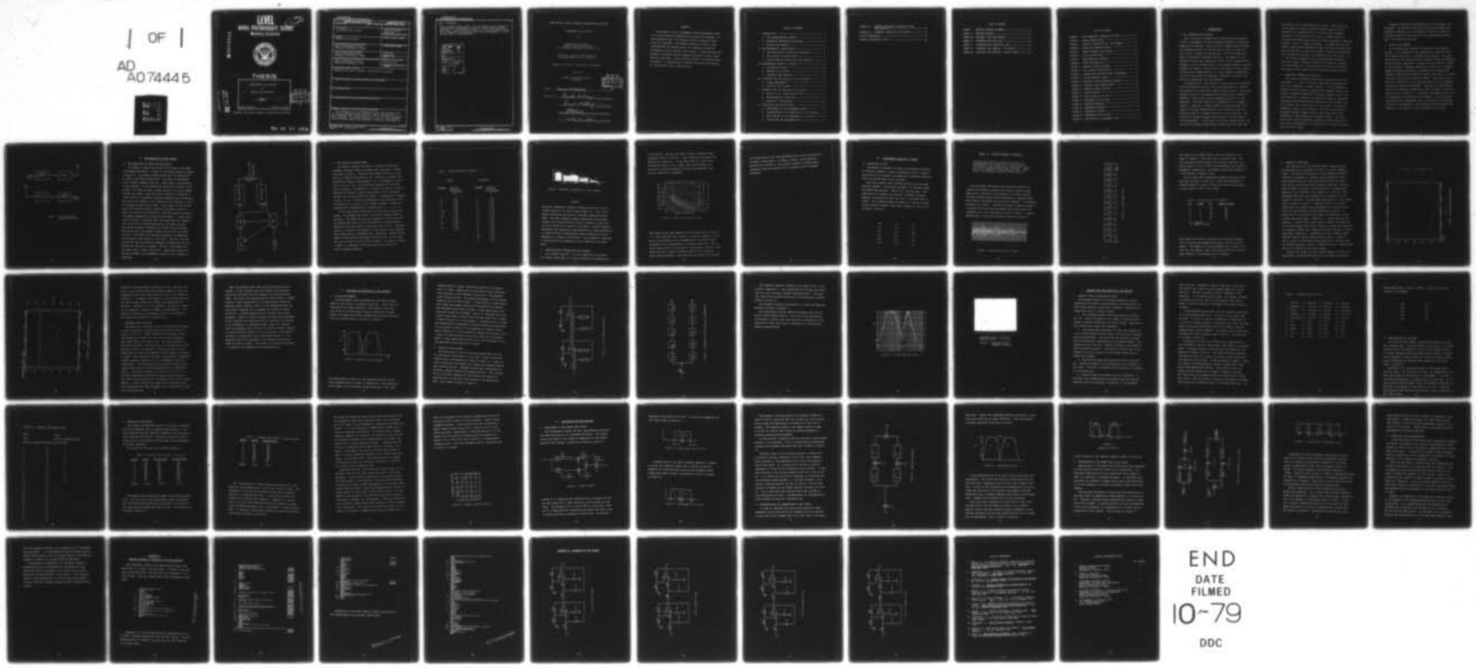
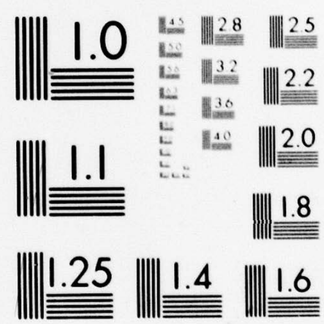
AD-A074 445 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 9/4
NARROWBAND VOICE SYSTEM.(U)
JUN 79 R R HITCHCOCK

UNCLASSIFIED NL

1 OF 1
AD
A074445



END
DATE
FILMED
10-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

LEVEL II

B.S.

NAVAL POSTGRADUATE SCHOOL

Monterey, California

9 Master's THESIS,

6 NARROWBAND VOICE SYSTEM.

by

10 Raymond Roy/Hitchcock

11 Jun[redacted]979 12 65p.

Thesis Advisor: Gerald D. Ewing

251 450

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
Narrowband Voice System

5. TYPE OF REPORT & PERIOD COVERED
Master's Thesis;
June 1979

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Raymond Roy Hitchcock

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Postgraduate School
Monterey, California 93940

12. REPORT DATE
June 1979

13. NUMBER OF PAGES
64

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)
Naval Postgraduate School
Monterey, California 93940

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The problem of how to accommodate more simultaneous users in a given bandwidth communications channel was examined. It was determined that voice signals could be reduced in bandwidth by 50 percent. Two speech formants, F1 and F3, were filtered from the speech signal and frequency translation performed to

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
(Page 1) S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Continued.

form a compact audio signal. At the receiver this process is reversed and the formants returned to their normal spectral positions. Intelligibility scores of 94 percent were obtained in a listening test conducted using the two formants F1 (250-650 Hz) and F3 (1950-2450 Hz).

NARROWBAND VOICE SYSTEM

by

Raymond Roy Hitchcock
Captain, United States Army
B.S., Virginia Military Institute, 1969

Submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
June 1979

Author Raymond R. Hitchcock

Approved by: Gerald D. Ewing
Thesis Advisor

Donald A. Stentz
Second Reader

D E Kirk
Chairman, Department of Electrical Engineering

William M. Tolles
Dean of Science and Engineering

ABSTRACT

The problem of how to accommodate more simultaneous users in a given bandwidth communications channel was examined. It was determined that voice signals could be reduced in bandwidth by 50 percent. Two speech formants, F1 and F3, were filtered from the speech signal and frequency translation performed to form a compact audio signal. At the receiver this process is reversed and the formants returned to their normal spectral positions. Intelligibility scores of 94 percent were obtained in a listening test conducted using the two formants F1 (250-650 Hz) and F3 (1950-2450 Hz).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. THE COMMUNICATIONS PROBLEM

One of the major problems facing communications engineers is how to accommodate the ever-increasing number of subscribers competing for services. Many of today's communications systems will become saturated in the near future; that is, they will not be able to accept new subscribers. This has already happened to the mobile telephone services. One aspect of this problem of particular importance to the military communications engineer is voice communications, the principal means of tactical command and control. The problem to be solved is how to afford reliable voice communications to a large number of subscribers given the limitations of bandwidth, frequency allocation, and the costs associated with new equipment.

Two approaches to the solution of this problem are apparent. The first approach requires that more frequencies or channels be acquired for a particular communications service. The second approach relies on the baseband voice signal being reduced in bandwidth, thus allowing more subscribers access to a given bandwidth. The first solution is impractical for a number of reasons. There are a limited number of frequency bands allocated to military communications. Any request for more frequencies would be in competition with civilian services. Additionally, adding frequency bands requires the acquisition of new equipment. In the case of satellite communications, the cost of an existing system cannot be amortized over less than

its design life of approximately six years. This means that no new satellite having increased channel capacity can be launched until replacement of an existing system becomes necessary. The second solution has the primary result of allowing more subscribers access to existing systems by means of external, "add on," speech processors. M. Weber reported that using a compact band modulation system two voice and three teletype channels were transmitted through a satellite transponder with a bandwidth designed to accommodate one voice channel. [1]

The solution approach chosen was bandwidth reduction of the voice signal. The major parameter assigned as a measure of effectiveness was intelligibility of the received signal. Voice quality and fidelity were considered secondary factors.

B. BANDWIDTH COMPRESSION TECHNIQUES

Four types of bandwidth compression techniques have been developed. [2] Time or frequency compression techniques exploit the redundancy of speech signals. In time compression methods the voice signal is sampled and redundant samples discarded. Frequency compression techniques select bands of the voice spectrum, discarding the remaining frequency components. Continuous analysis-synthesis techniques do not transmit the voice signal, but a description of the signal in the form of parametric analog control signals. Discrete sound analysis-synthesis systems transmit binary code groups identifying fundamental sounds comprising the speech signal. The last, sound group analysis-synthesis methods, transmit groups of binary codes corresponding to a selected set of words or phrases from the voice signal.

Frequency compression was selected as the technique to be developed. An analysis of the four techniques revealed that frequency compression of the baseband voice signal would be the most compatible with existing voice communications equipment, both analog and digital.

C. SOLVING THE PROBLEM

Voice signals were analyzed to determine their spectral content. This information was considered as a description of the transmitted signal in a communications system. The characteristics of the human ear were analyzed in order to determine which frequency components of speech are best received. Based on this analysis, frequency bands were selected and a bandpass speech processor constructed. Processed voice signals and standard communications bandwidth signals were used in an intelligibility test and the results compared. After verification that the intelligibility criterion had been met, the frequency translation system was developed. The result of this process was a voice bandwidth compression system, as shown in Figure 1. Analog voice is reduced to a contiguous narrowband signal prior to transmission through existing equipment. At the output of the receiver the compressed audio is restored to its original position in the spectrum. The following sections provide the details of the analysis of the narrowband voice system.

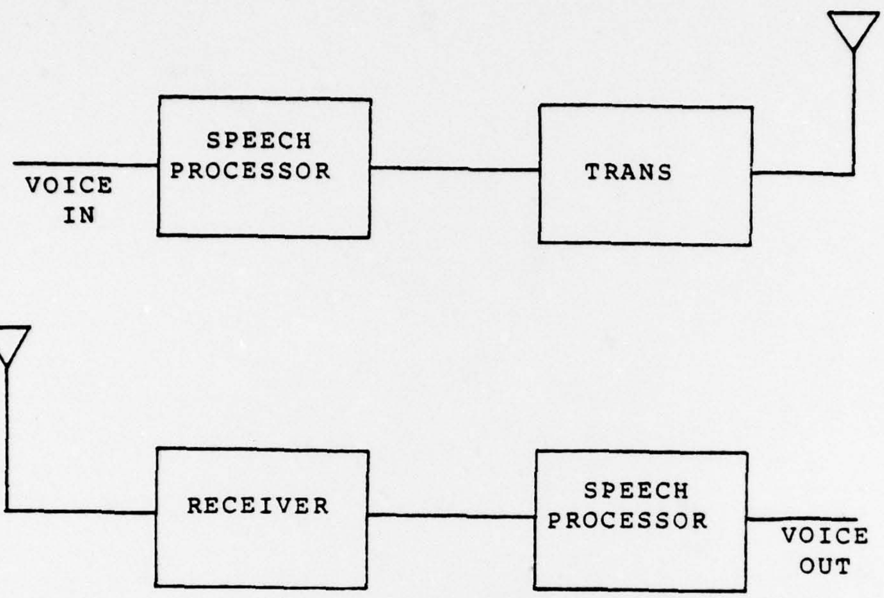


FIGURE 1 VOICE BANDWIDTH COMPRESSION SYSTEM

II. THE MECHANISM OF HUMAN SPEECH

A. THE PRODUCTION OF SOUNDS AND MODULATION

The concept of speech was found to be similar to an amplitude modulation system. A model of the speech process is shown in Figure 2. The speech process begins with the formation of an idea to be orally communicated. The idea is transformed into a complex language code set. These codes, corresponding to the phonemes, are the modulating and control signals in the system. The lungs provide a steady stream of air which passes through the larynx. When the larynx vibrates, it produces a signal having a fundamental frequency determined by two factors, the size of the opening in the larynx and the intensity of the air stream provided by the lungs. This signal is rich in harmonics and normally covers a spectrum from 80Hz to 8000 Hz. When this carrier is produced by a vibrating larynx, the resulting speech components are termed voiced sounds. A second category of carrier is produced when the larynx does not vibrate, that is by the air stream passing over the teeth and lips. The resulting speech components are termed unvoiced sounds. The carrier wave then passes through both the oral and nasal cavities. The nasal cavities serve primarily as resonators, giving quality and fidelity to the voice. The oral cavities contain the four articulators necessary for modulation of the carrier and formation of speech. The four articulators are the lips, teeth, tongue, and palate. These work together to generate phonemes, the fundamental sounds of which speech is comprised.

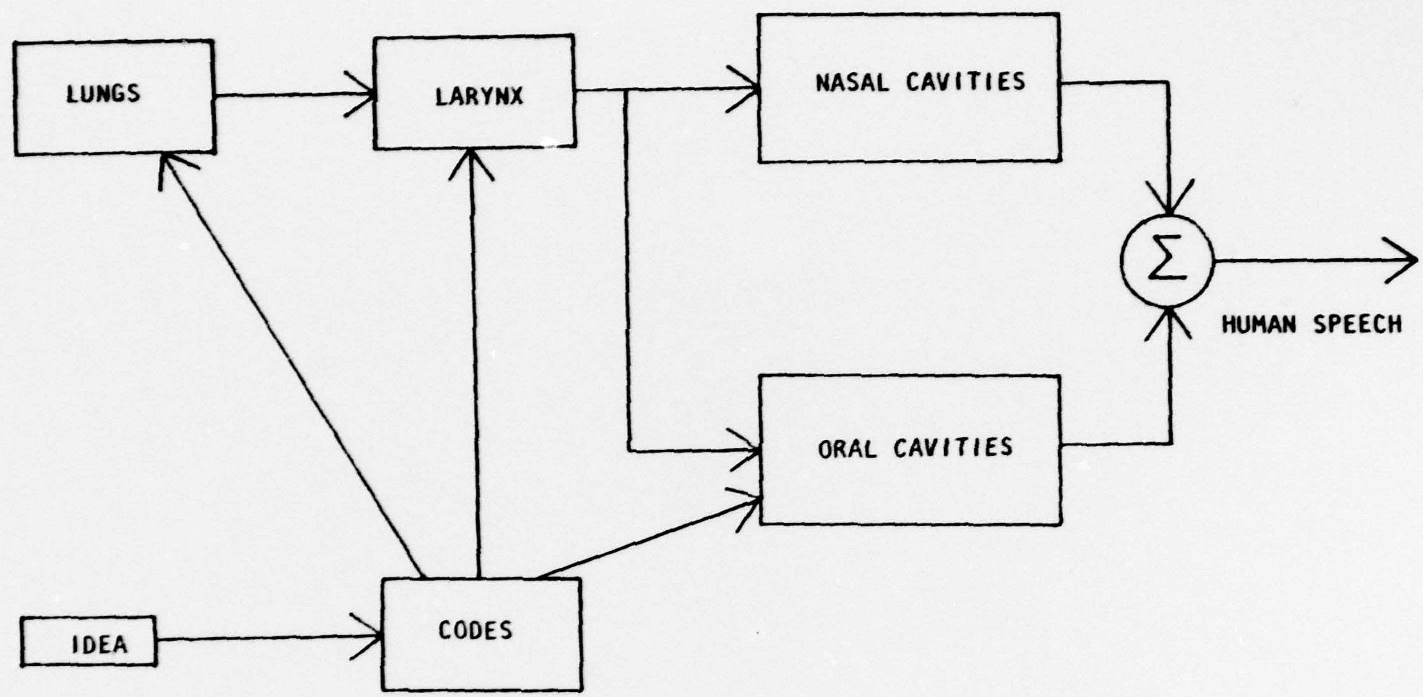


FIGURE 2 SPEECH PROCESS MODEL

B. THE FACTORS OF HUMAN SPEECH

The English language was found to consist of forty-two phonemes; eighteen vowels and twenty-four consonants. These are shown in Table I. The four articulators interact to produce five types of articulation [3]. Plosives, or stops, are produced by stopping the passage of air. An example of a plosive is the p in pop. Fricatives are produced by narrowing the air passage. An example of a fricative is the th in their. Laterals are formed by closing the middle line of the mouth allowing air to pass around the tongue. An example of a lateral is the l in let. Trills are produced by rapid vibration of an articulator as a rolled r in the German language. Vowels are produced by an unobstructed air passage as a, e, i, o, or u.

If the larynx were constrained to vibrate at a single frequency, the phonemes and, hence, the speech produced would be monotone. The codes changing the frequency of the carrier during speech were found to stem from emotional and personality factors giving each person a distinctive voice print. Although there is a variation in frequency of the carrier during speech, four common carrier frequencies were identified. These are the result of the physical resonance characteristics of the nasal and oral cavities, and are called formants. As shown in Figure 3 there is a decrease in the voice power associated with higher frequency formants.

TABLE 1 ENGLISH LANGUAGE PHONEMES

VOWELS

PHONEME	RELATIVE FREQUENCY OF OCCURENCE (%)
I	8.53
a	4.63
æ	3.95
ɛ	3.44
ə	2.81
^	2.33
i	2.12
e,ei	1.84
u	1.60
ai	1.59
oU	1.30
ɔ	1.26
U	0.69
aU	0.59
a	0.49
o	0.33
ju	0.31
	0.09

CONSONANTS

PHONEME	RELATIVE FREQUENCY OF OCCURENCE (%)
n	7.24
t	7.13
r	6.88
s	4.55
d	4.31
l	3.74
ð	3.43
z	2.97
m	2.78
k	2.71
v	2.28
w	2.08
p	2.04
f	1.84
h	1.81
b	1.81
ŋ	0.96
ʃ	0.82
g	0.74
j	0.60
tʃ	0.52
dʒ	0.44
θ	0.37
ʒ	0.05

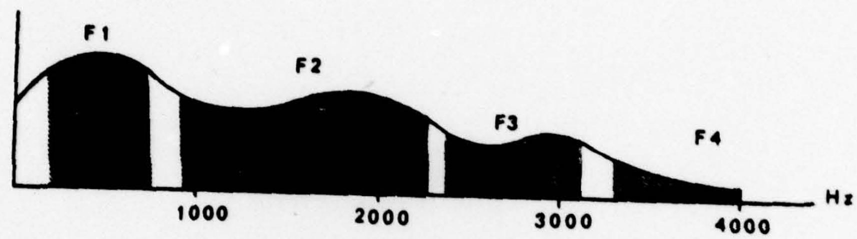


FIGURE 3 SPECTRAL DISTRIBUTION OF THE FORMANTS

Figure 3

The major information content of speech was found to fall in formant areas F1, F2, and F3 below 3000Hz [2]. Very little speech information was found in frequency components below 200Hz. The Bell Telephone Company determined that suitable fidelity and intelligibility could be attained by band limiting voice signals to a range of 200-3200Hz. Further research determined that acceptable intelligiblity with some degradation in fidelity could be obtained by further reducing the voice bandwidth to a range of 200-2500Hz. This standard is employed in most military and commercial voice communications equipment.

C. INTELLIGIBILITY FACTORS AND THE RECEIVER

The ultimate receiver in a voice communications system is a human, whose ears act as the collectors and demodulators

of the signal. The ear was found to have a frequency characteristic shown in Figure 4. This figure was extracted from Fletcher [4, page 135]. It was seen from this figure that the hearing acuity of the tested group varied widely as a function of required intensity at any given frequency, but not as a function of frequency.

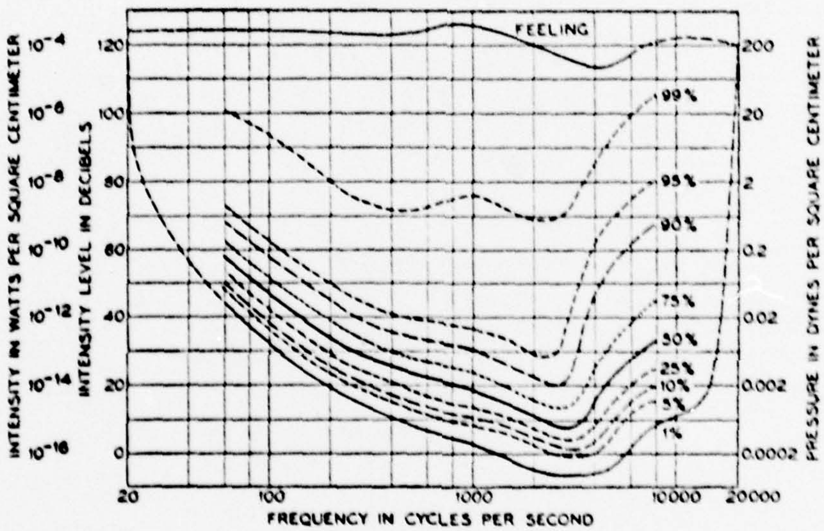


FIGURE 4 HUMAN EAR RESPONSE CURVES

When these curves were compared to the formant areas in Figure 3, it was found that the increase in required intensity in the ear at low frequencies was accommodated by the higher power allocated to the frequencies by the speech mechanism. The relation holds true to a frequency of approximately 2000 Hz, past which required intensity increases rapidly, but available speech power decreases. The human ear was found to be matched

to the portions of the voice spectrum which contain the majority of speech intelligence. In order to obtain a more detailed quantitative analysis of the spectral context of human speech, recorded voices were analyzed by a discrete Fourier transform technique.

III. EXPERIMENTAL ANALYSIS OF SPEECH

A. COLLECTION OF DATA

The method of analysis of actual voice signals consisted of recording speakers' voices, performing analog to digital conversion of the signals, and computing a frequency spectrum from the discrete data points obtained.

Four male speakers were selected and each read the same selected message. Voices were recorded on a Panasonic model RQ-212DKS tape recorder. TDK SA C-90 tape was used. This combination of recorder and tape had a measured frequency response characteristic which was flat in the range 60Hz to 8150Hz. The recorded message consisted of a series of three letter phonetic alphabet groups and a passage of plain text, as shown in Table II.

BCD	QVV	CRP
XFZ	PVM	XSX
KPC	TDK	YTW
MNX	GNI	AUQ
DFB	BJF	DWE
EHO	JLK	IYJ

TABLE II. SELECTED MESSAGE (CONTINUED)

STUDENTS ARE REMINDED TO SUBMIT ALL CLAIMS FOR REIMBURSEMENT UPON COMPLETION OF TRAINING NOT LATER THAN 30 DAYS AFTER CLOSE OF THE FISCAL YEAR (30 SEPT) IN WHICH EXPENSES WERE INCURRED. CLAIMS ARE TO BE SUBMITTED TO USA ADMINCEN, ATTN: ATZI-PA-RM, FORT BENJAMIN HARRISON, INDIANA 46216.

Once recorded, the analog voice was band limited to a range of 200-3200Hz by passing the signals through a band-pass filter consisting of two Kronhite 3321 audio filters. The band limited signals were processed by a Commodore 500C-Data General 9300 hybrid analog-digital computer. The sampling frequency was 7600 Hz. Data was processed through an analog to digital converter which provided 14 bit word output corresponding to a full scale analog signal of ±5Vpp. A typical sampled input to the A/D converter is shown in Figure 5.

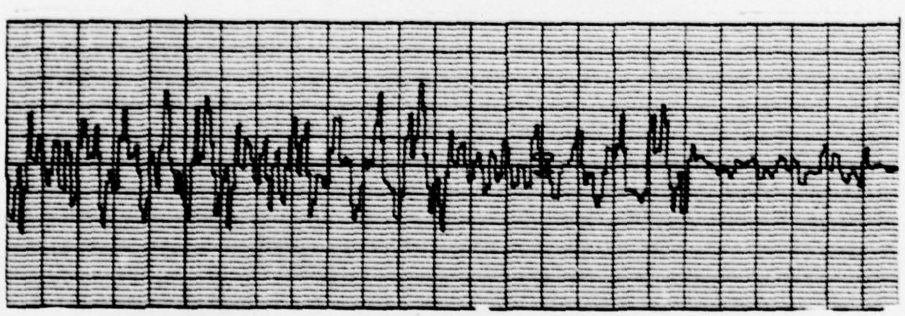


FIGURE 5 SAMPLED ANALOG VOICE SIGNAL

-1 FS	$+2^{-1}$ FS	$+2^{-2}$ FS	$+2^{-3}$ FS	$+2^{-4}$ FS	$+2^{-5}$ FS	$+2^{-6}$ FS	$+2^{-7}$ FS	$+2^{-8}$ FS	$+2^{-9}$ FS	$+2^{-10}$ FS	$+2^{-11}$ FS	$+2^{-12}$ FS	$+2^{-13}$ FS
BIT 13													BIT 0

FIGURE 6 14 BIT OUTPUT WORD

The format of the output word of the A/D converter is as shown in Figure 6. Note that only 14 bits are used. The word represents analog signals in the range ±5Vpp by a converted range of -5V to +4.99V. This difference was ignored in subsequent calculations. The digital words were recorded in 7 track format on magnetic tape.

Files of data were organized by speaker and type of message text as shown in Table III. Each file consisted of a variable number of logical records units each record having a fixed length of 4096 14 bit words.

TABLE 3 RECORDED FILES AND MESSAGES

FILE	SPEAKER	TEXT	NUMBER OF RECORDS
1	1	PA	70
2	1	PT	35
3	2	PA	105
4	2	PT	43
5	3	PA	76
6	3	PT	46
7	4	PT	137

PT - PLAIN TEXT
PA - PHONETIC ALPHABET

The number of records is different for different speakers. This is because the speakers have widely varying rates of speech. The emphasis on file management was in order to insure that the samples used in the DFT process were from the same elements of the message for all speakers.

B. ANALYSIS OF THE DATA

The calculation of the discrete fourier components was accomplished on the IBM 360/67 at the W. R. Church Computer Center. The detailed discussion of this computer program is contained in Appendix A. Analysis was organized by files. Three records from each file were analyzed with the second 1024 words in each record being used as input data. The program provided, as output, the amplitudes of 512 unambiguous harmonics with a spectral line interval of 7.3241 Hz. This frequency resolution was determined to be adequate for the purpose of analysis. The mean amplitude spectrum of the tested group is shown in Figure 7. Comparing Figure 3 to Figure 7 showed that the formant areas correspond well enough to conclude that the formants were experimentally verified. By making use of the relation that power is proportional to the voltage squared and $P_{dB}(I) = 20LOG(V(I))$, a power spectrum was computed. A graph of the power spectral density was obtained for the same sample sets used to compute the amplitude spectrum. The computer program which computes the power spectrum is also contained in Appendix A. Examination of the different power spectral density graphs showed distinct peaks in formant F1 and F3, while F2 contained multiple peaks with a pattern varying between speakers. Figure 8, taken from file 6, demonstrates this and is typical of the other graphs.

Although the reason for this random behavior in F2 was not discovered, it was conjectured that this formant contributes

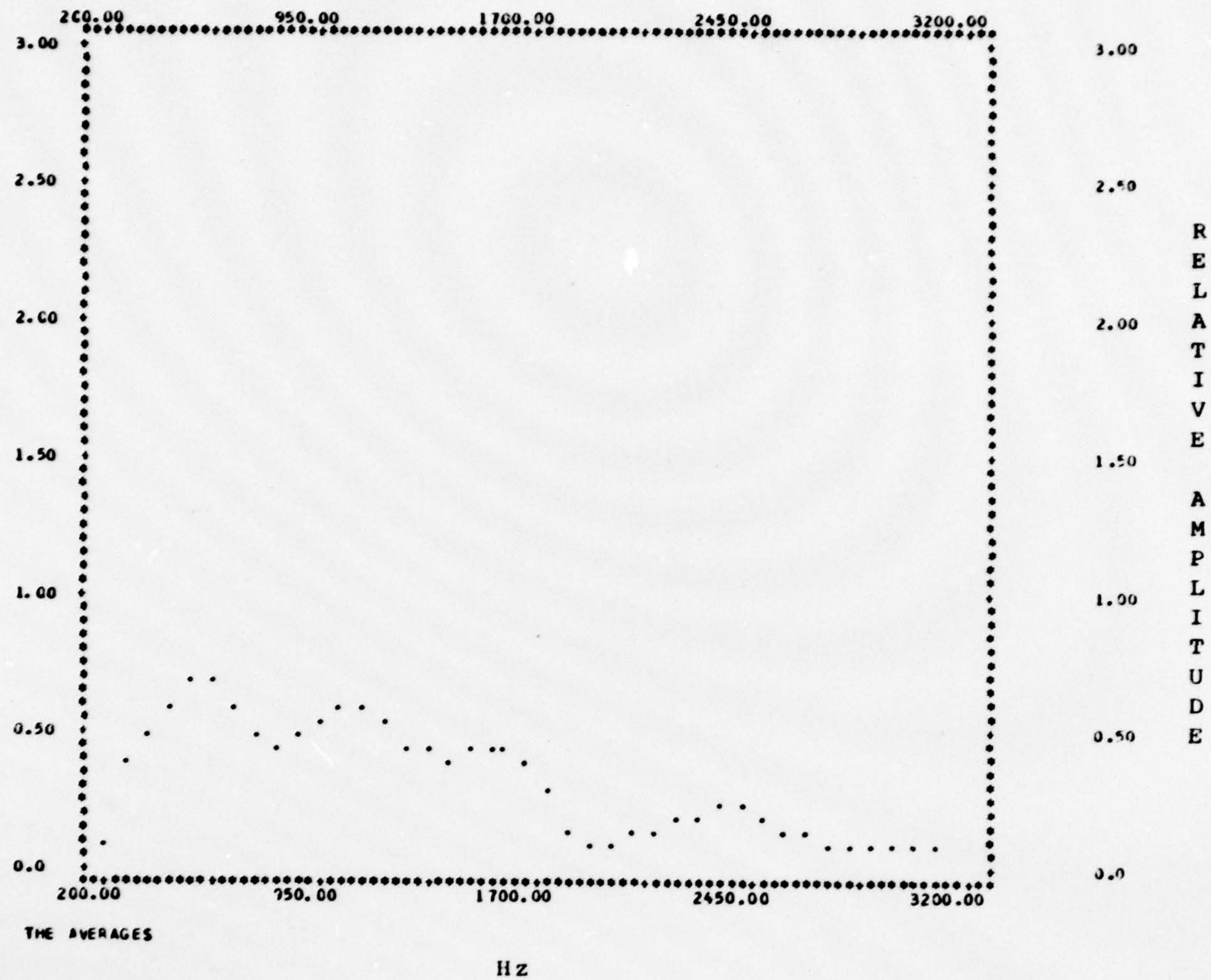


FIGURE 7 MEAN AMPLITUDE SPECTRUM

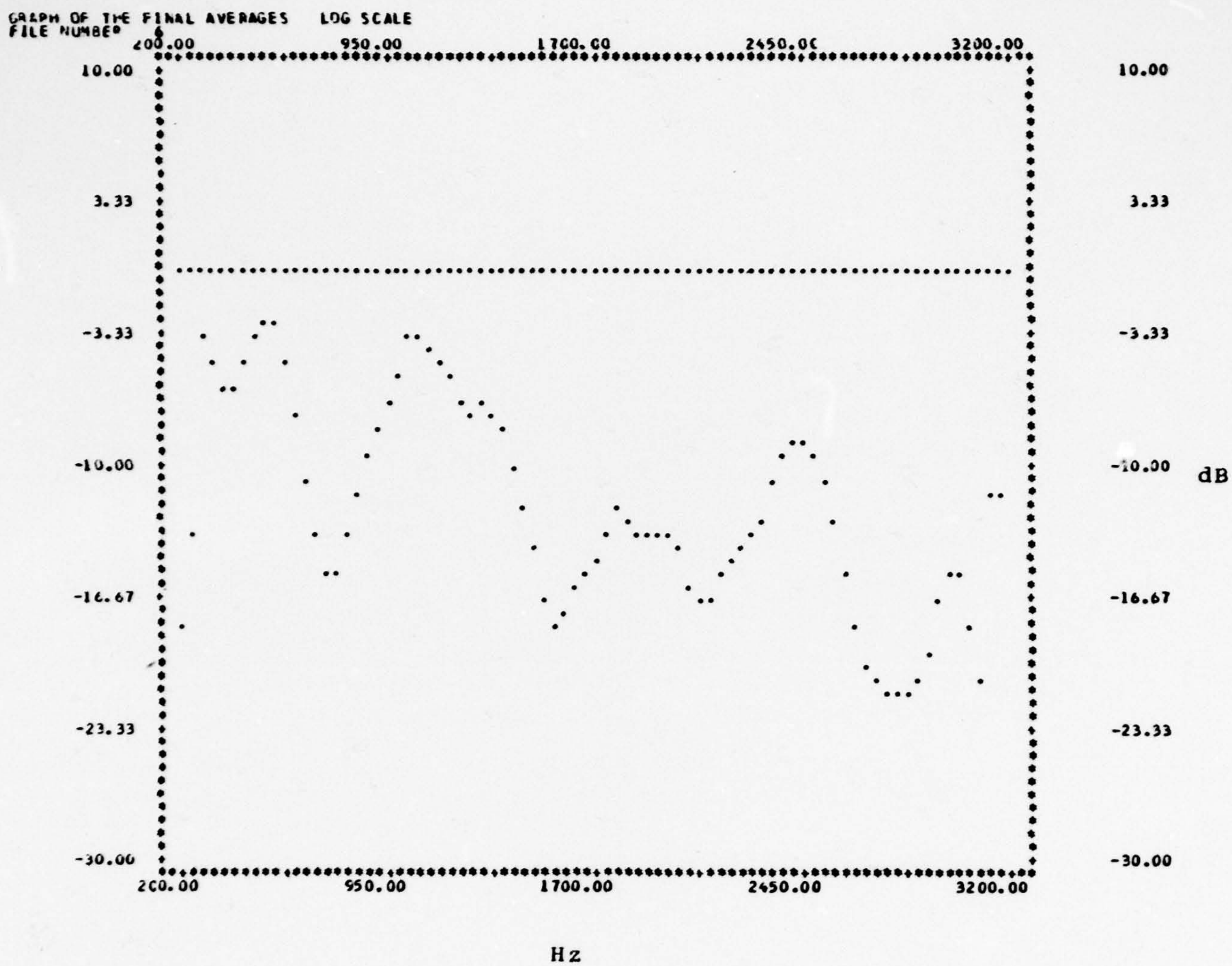


FIGURE 8 POWER SPECTRUM FOR FILE 6

strongly to the personality factors of voice. The only variables in the analysis were the different speakers, hence the possibility that those identifying factors of voice appear in formant F2. In summary, the result of this analysis was that the three formant areas below 3200Hz were experimentally verified. Formants F1 and F3 displayed a similarity in shape for all speakers, whereas F2 showed a random behavior. The problem of selecting the frequency bands to be isolated for the compression scheme was considered next.

C. FREQUENCY BAND SELECTION

The question of how much of the voice spectrum was needed in order to maintain intelligiblity was studied by Karl D. Kryter in 1957 [5]. Kryter examined intelligibility scores when speech was divided into one, two or three passbands, each 500Hz wide at the -6dB points. The data presented in that paper provided a starting point for determination of the frequency bands to be selected. Kryter concluded that any system should include one passband centered around a frequency of 500Hz. This conclusion has been corroborated by subsequent research showing formant F1 to be a fundamental carrier of intelligence. With sufficiently high signal-to-noise ratios (20dB) intelligibility can be maintained while transmitting only F1. This was not found to be true for F2 or F3 in a listening experiment conducted by passing speech through a band pass filter consisting of two Kronhite 3321 audio filters. However, it was concluded that under normal signal-to-noise ratio conditions one 500Hz passband was not sufficient to maintain intelligibility.

When two passbands were used intelligibility scores increased. Kryter reported that the highest intelligibility scores were obtained when one passband was centered around 500Hz. The second was centered either around 1500Hz or 2500Hz. Comparing these frequencies to the experimentally obtained spectrum graphs, Figures 7 and 8, it was determined that the true center frequency for a passband for formant F3 should be 2200 Hz. Although Kryter did not conclude which center frequency, 1500Hz or 2500Hz, provided the best second passband, he did conclude that these were the two choices to be studied in any subsequent two passband system. Weber [1] reported that two passbands, 250 to 800 Hz and 1900 to 2450Hz, resulted in intelligibility scores of 90 to 100 percent. The decision was made to concentrate further development based on a two passband system with passbands in the frequency ranges 250 - 800Hz and 1900 to 2450Hz. The design of the filters necessary to accomplish this separation was considered next.

IV. FILTERING AND SELECTION OF THE SPECTRUM

A. FILTER REQUIREMENTS

Two requirements were formulated for the audio filters based on the concept of frequency selection. First, the filters had to be highly selective. The slope of the filter skirts had to be sufficiently steep to allow the two pass bands to be separated without leakage from the adjacent formant due to undue overlap, as shown in Figure 9.

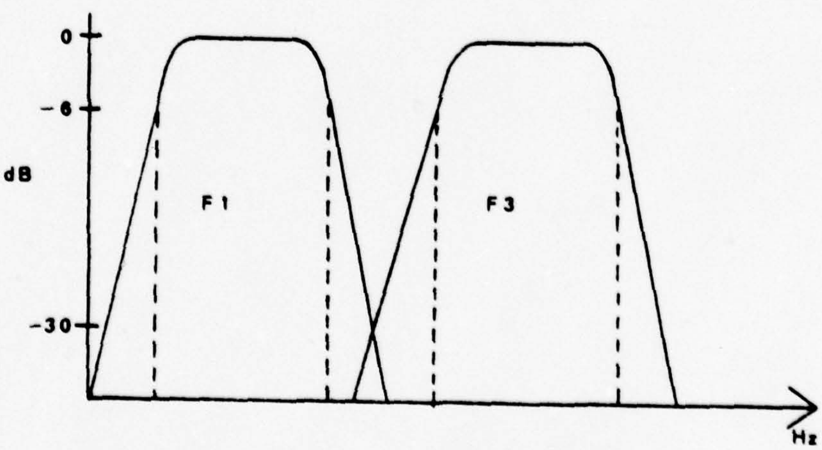


FIGURE 9 REQUIRED FILTER RESPONSE

The characteristic chosen for the outside skirts was an 8th order response having a slope of -48dB/octive. The characteristic chosen for the adjacent inside skirts was a 12th order

response having a slope -72dB/octive providing a crossover point of -30dB. Examination of the various filter configurations resulted in the Chebyshev filter with .1 dB passband ripple being selected. The second requirement was for minimum distortion due to filter delay. Filters having a higher passband ripple and steeper skirts were considered and rejected. The Chebyshev filter does not have a linear phase delay characteristic across the passband. Harris [6] reported that with higher ripple factor filters this nonlinear phase delay caused a severe distortion of voice signals. Crossover point of -30dB was experimentally determined to provide sufficient suppression of the adjacent formant for the purposes of listening tests. When listening to a voice recording, another voice signal units a level of -30dB relative to the first could not be detected. Based on these specifications the filters were designed utilizing operational amplifier active filters.

B. DESIGN OF THE FILTERS

The realization of the filters was accomplished using LM 741 operational amplifiers in a 4th order voltage controlled-voltage source configuration as shown in Figure 10. These 4th order filters were configured into lowpass and highpass modular basic building blocks. Bandpass filters were constructed by cascading lowpass modules with highpass modules. The required 8th or 12th order response for each module was obtained by cascading two or three 4th order modules of the appropriate type. This scheme is shown in Figure 11.

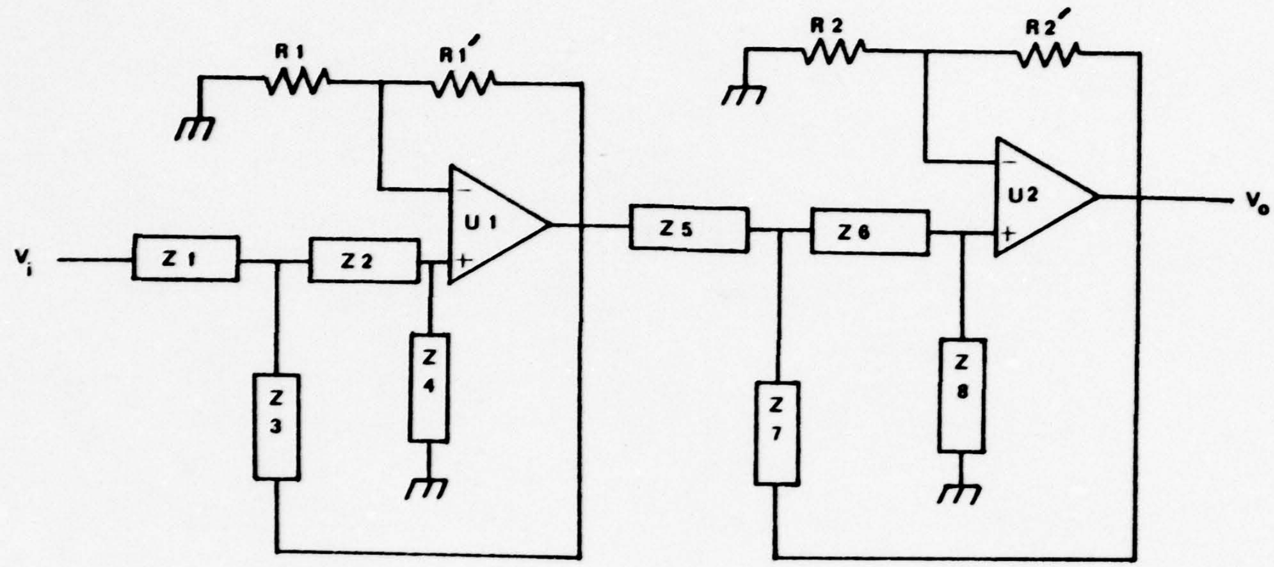


FIGURE 10 VCVS FILTER CONFIGURATION

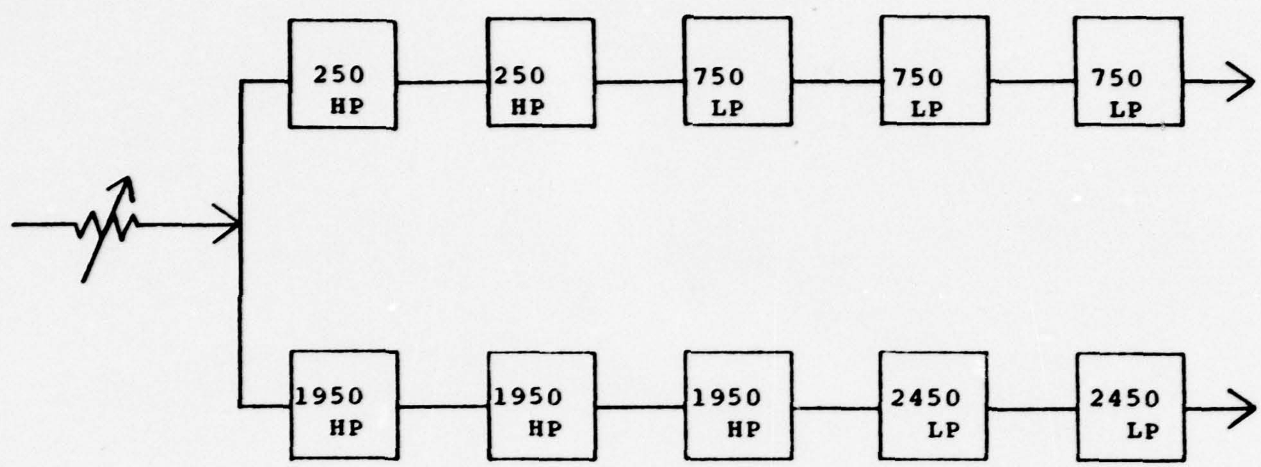


FIGURE 11 LOWPASS AND HIGHPASS MODULE ARRANGEMENT

The complete schematic diagram of the audio filter is contained in Appendix B. Once constructed the filters were evaluated for their frequency response characteristics. The measured test data are shown plotted with the theoretical response curves in Figure 12.

The frequency response as measured on a MICRO FFT spectrum analyzer is shown in Figure 13.

The experimental results compared favorably with the objective design specifications. With the filters operational, the baseband listening test was conducted to determine if the frequencies chosen would allow transmission of intelligible speech as hypothesized.

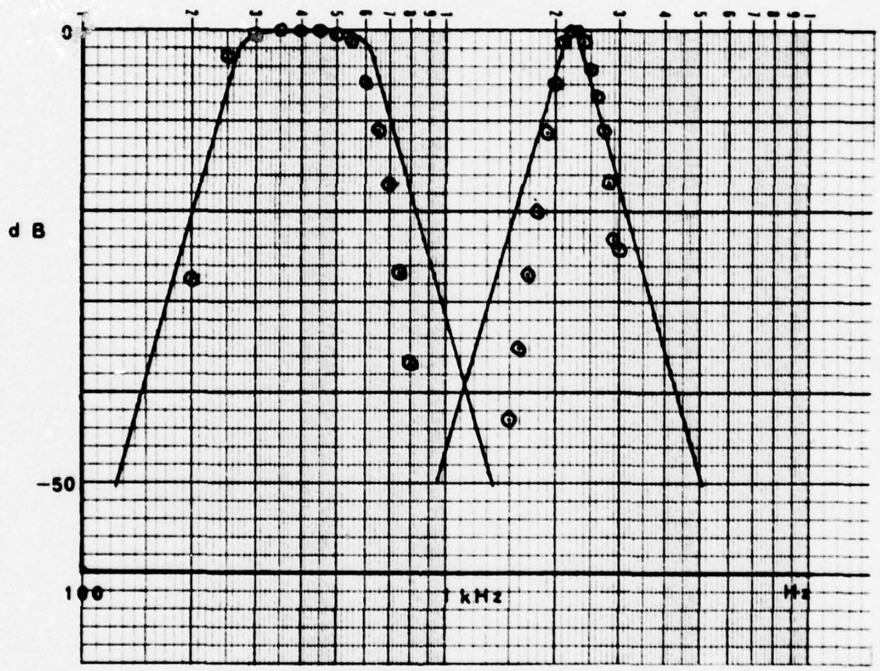
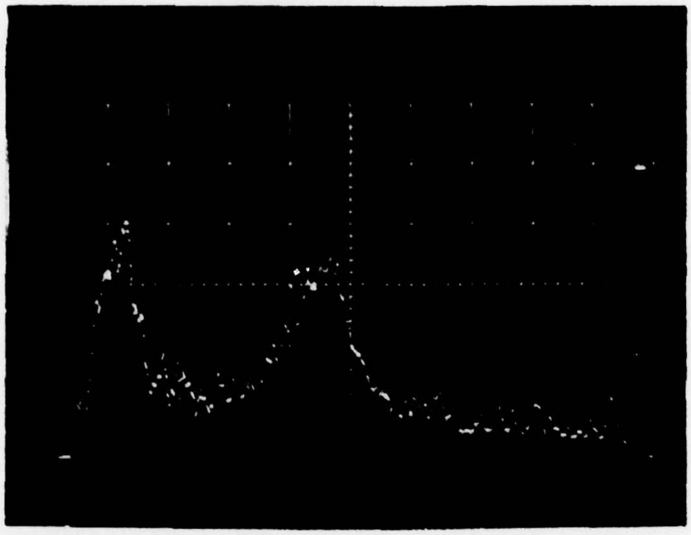


FIGURE 12 FILTER RESPONSE CURVES



HORIZONTAL SCALE 0-5000 Hz
VERTICAL SCALE 0- -60 dB

FIGURE 13 MEASURED FILTER FREQUENCY RESPONSE

V. BASEBAND TEST AND ANALYSIS OF THE RESULTS

A. DESIGN OF THE INTELLIGIBILITY TEST

The absolute measure of the effectiveness of a voice communications system is determined by a listening test in which human subjects record their responses. There are four recognized types of listening tests:

1. The nonsense syllable test uses words such as deeg and zak. Scoring is based on the number of vowels and consonants correctly determined from each syllable. Each syllable contains both vowels and consonants.

2. The phonetically balanced (PB) word test consists of a set of words taken from the twenty lists of Harvard Phonetically Balanced Words. The words are one syllable in length and are phonetically representative of everyday speech. Two examples are mud and sled. Scoring is based on the number of words correctly recorded. A variation of the PB list is the Spondee list which consists of two syllable words such as hotdog and airplane.

3. The modified rhyme test uses words such as rang and pit. Scoring is based on the recognition of the phonemes in each word. This type is accepted by the military as a measure of intelligibility.

4. Sentence tests are divided into two categories. In the first, the listener hears the sentence and then answers questions about the sentence to determine if he understood

what was said. For example, hearing "John went to the store for bread," the listener would be asked "where did John go, and why?" Scording is based on correct responses to the questions. In the second type of test, the listener is asked to identify key words, which in this case are John, store, bread. Scoring is based on the number of key words correctly recorded.

It was decided that the basis for the listening test would be a PB and Spondee Word List. A total of forty words were selected. In order to provide a standard for comparison and to give the listener practice in performing the test, the first twenty words were bandlimited from 200 to 2500Hz, standard communications channel bandwith. The second twenty words were filtered into the two formant areas F1 and F3. The word list is as shown in Table IV.

The majority of tactical military voice communications are conducted by means of three letter code groups of encrypted words. The basic document which provides the encryption and decryption tables is the SOI. The current Standing Operating Instructions (SOI) will contain an encoding table translating key words into code groups. A typical message "Company C needs 100 rounds of TOW Ammunication" would be transmitted in the format "YKX BMN WBD CYT AAK PMC." Each letter of the code groups would be spoken as its phonetic alphabet character. The first word of the message would be transmitted as "Yankee Kilo X-ray." For this reason a series of ten phonetic alphabet code groups was included as words in the narrow band test. The

TABLE 4 LISTENING TEST WORD LIST

1.	Airplane	11.	Baseball	21.	Daybreak	31.	Cupcake
2.	Woodchuck	12.	Doorstep	22.	Shotgun	32.	Workshop
3.	Outlaw	13.	Lifeboat	23.	Mishap	33.	Greyhound
4.	Headlight	14.	Farewell	24.	Pinball	34.	Railroad
5.	Blackboard	15.	Wildcat	25.	Hotdog	35.	Sundown
6.	Elk	16.	Tip	26.	Price	36.	Owe
7.	Grape	17.	Elm	27.	Muff	37.	Moth
8.	Smash	18.	Mode	28.	Fell	38.	Hack
9.	For	19.	Soap	29.	Scuff	39.	Gem
10.	Pug	20.	Nag	30.	Thin	40.	Sled

code groups chosen, listed in Table V, contain all 26 characters of the alphabet.

RHZ	YTP
MKP	UXS
JMA	TDO
BFN	VEF
GCK	LQI

B. ADMINISTRATION OF THE TEST

Five male and five female listeners were selected to take the listening test. The group ranged in age from 31 to 34 years old and none of the subjects claimed to have abnormal hearing or current temporary hearing problems such as colds. The subjects listened to a tape recording of the test through communications earphones and recorded their responses on the form shown in Figure 14.

Instructions to the subjects were to write down exactly what was heard. If a word was not clear, for example bar or car, their best guess of the word was to be recorded. If no word could be distinguished, no response was to be recorded. All female subjects were furnished a copy of the phonetic alphabet to be used as a reference in part three of the test. Since the male subjects were familiar with the PA, it was not necessary to give them a list.

FIGURE 14 LISTENING TEST ANSWER SHEET

NAME: RANK:

AGE: HEARING PROBLEMS:YES/NO

SERVICE: USA/USN/USMC/USAF/USCG/CIVILIAN

1.	21.	1.
2.	22.	2.
3.	23.	3.
4.	24.	4.
5.	25.	5.
6.	26.	6.
7.	27.	7.
8.	28.	8.
9.	29.	9.
10.	30.	10.
11.	31.	
12.	32.	
13.	33.	
14.	34.	
15.	35.	
16.	36.	
17.	37.	
18.	38.	
19.	39.	
20.	40.	

C. ANALYSIS OF THE RESULTS

The answers recorded were scored in two ways, consistent with the recognized tests of speech intelligibility. The first scoring technique used was consistent with the phonetically balanced word method. Words were scored as right or wrong and the intelligibility score taken as the percentage of correct responses.

The results of the test are as shown in Table VI.

TABLE 6 LISTENING TEST RESULTS - PB SCORING

SUBJECT	COMM BW	PROCESSED WORDS	PROCESSED GROUPS
1	100	85	100
2	100	95	100
3	95	90	100
4	100	95	100
5	100	95	100
6	100	100	96.87
7	100	100	100
8	100	100	100
9	100	90	100
10	95	85	93.33
AVERAGE	98.99 %	93.49 %	99.0 %

The second scoring method was based on the modified rhyme test. One point was given each phoneme correctly recorded. Only the one syllable words were considered in the analysis and the phonetic alphabet test was not used. The results are as shown in Table VII.

TABLE 7 LISTENING TEST RESULTS - RHYME SCORING

SUBJECT	COMM BW	PROCESSED WORDS
1	100	85
2	95	100
3	95	90
4	100	100
5	100	90
6	100	100
7	100	100
8	100	100
9	100	90
10	95	85
AVERAGE	98.5 %	94.0 %

The interpretation of these scores was developed for two areas of voice communication - plain text and code groups. The redundancy of the English language has been well documented. Fletcher [4] reported that only 737 root words form over 96 percent of an 80,000 word vocabulary. Essentially, recognition of only the root words would allow the listener to understand the spoken sentences with a high degree of intelligibility. Olsen [7] reported a relationship between syllable and sentence intelligibility shown in Figure 15.

The reason for sentence intelligiblity being much higher than the syllable or word intelligibility stems from two factors. The first factor is the redundancy of speech as previously discussed. It is not necessary to understand the whole word in order to understand the meaning. For example stop, stopper, stopping, stopped, and stops all have stop as a root word and the conceptual meaning of all these words is the same. The second factor is the concept of context. A sentence may be correctly interpreted even though words are missing or distorted. For example, the phase "a rolling stem gravels no mash," is easily understood because of its context and familiarity. This effect plays a very important part in speech intelligibility when the functional vocabulary is limited, as in most military communications.

Characteristically, the scores from isolated word list or syllable tests will be lower than scores from a test using sentences. Since no clear quantitative correlation between the tests exists, it was decided to use the scores obtained, realizing that actual system performance with sentences transmitted could be higher. Code group communications consist of words similar to those used in the listening test. For this reason it was decided to use the intelligibility scores for the phonetic alphabet test as obtained. Sentence intelligibility scores for other systems were examined in order to establish a reference. Toll quality telephone was found to have a score of 95 percent. Communications quality systems have a

score of 90 percent, while marginal communications could be maintained with a score of 70 to 80 percent. These figures represent averages. The conclusion was that satisfactory speech communications require only a small portion of the speech spectrum. Specifically, intelligibility of 93 to 99 percent was obtained while transmitting two passbands of speech, 250 to 750 Hz and 1950 to 2450 Hz, in experimental listening tests. With the hypothesis proven, the system was finalized in concept.

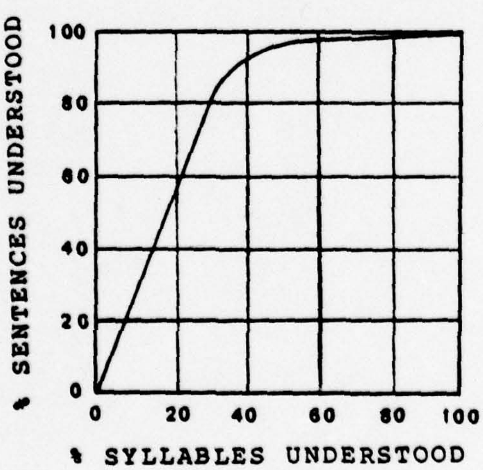


FIGURE 15 SENTENCE INTELLIGIBILITY

VI. CONCLUSIONS AND APPLICATIONS

A. DEVELOPMENT OF THE NARROW BAND SYSTEM

The two-passband concept had been experimentally verified and specifications for the passbands developed. The system design was based on the concept of separation of the speech spectrum and frequency translation as shown in Figure 16.

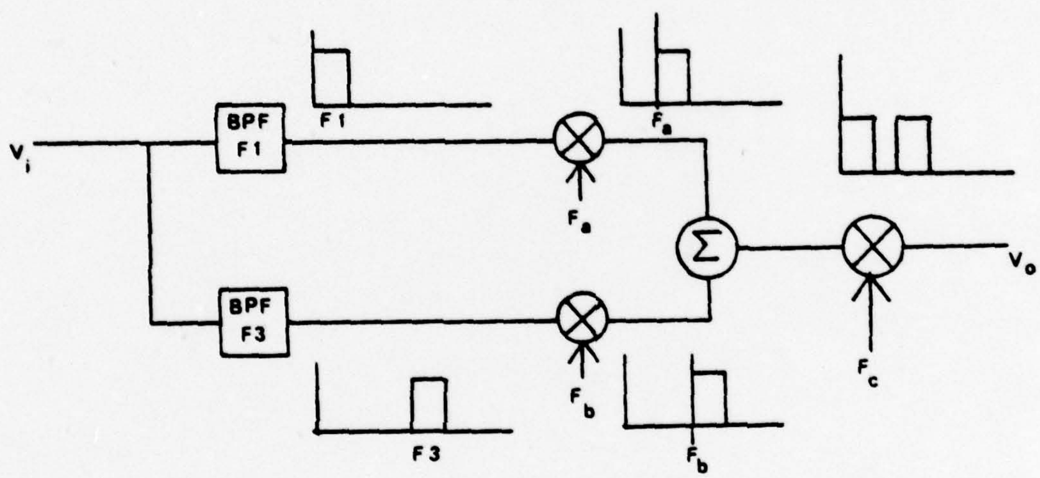


FIGURE 16 SYSTEM CONCEPT

Formant F1 is separated and translated up in frequency by mixing the signal with a 50kHz carrier F_A and filtering the sum terms. The spectrum of F1 is now 50.250 to 50.750 kHz. Formant F3 is separated and uptranslated by mixing the signal with a carrier F_B having a frequency of 48.850 kHz. F3 now has a

spectrum of 50.8 kHz to 51.3 kHz. F1 and F3 are summed giving the signal shown in Figure 17.

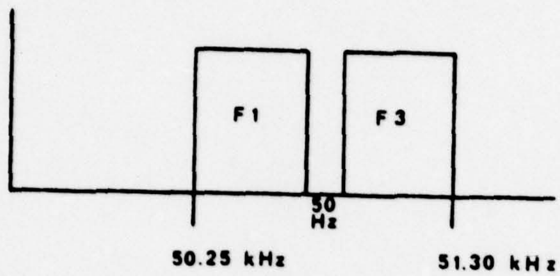


FIGURE 17 VOICE SIGNAL AT 50 kHz IF

A compact band in the audio frequency spectrum is formed by mixing the composite signal with a carrier F_c having a frequency of 50.000 kHz and filtering the difference terms. A 50Hz guard band has been left between F1 and F3, as shown in Figure 18.

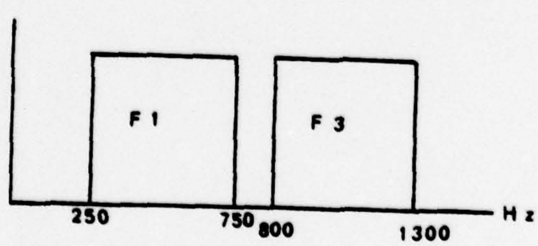


FIGURE 18 NARROWBAND VOICE SIGNAL

The purpose of the guard band is to provide a buffer between F1 and F3, realizing that the filters that will be used are not ideal and some mixing of signals will occur due to overlap. The frequency range of the compact signal is seen to be 250 to 1300 Hz, well within the system bandwidth of existing communications equipment.

At the receiver, formant F3 must be returned to the original position in the spectrum. This is accomplished by essentially reversing the sequence discussed above and is shown in Figure 19.

The audio output of the existing receiver is mixed with a carrier F_C having a frequency of 50.0 kHz. The composite audio spectrum is then separated into F1 and F3 by means of the filters shown. F1 is mixed with a carrier F_A having a frequency of 50 kHz and the difference terms retained. F1 has been returned to its original spectrum position of 250 to 750 Hz. F3 is mixed with F_B having a frequency of 48.850 kHz and the difference terms retained. F3 has been returned to its original spectrum position of 1950 to 2450 Hz. The two components are summed and the output spectrum consists of F1 and F3. This is the same audio spectrum which was utilized in the intelligibility tests. Considerations for implementation using nonideal filters was considered next.

B. CONSIDERATIONS FOR CONSTRUCTING OF THE SYSTEM

In order to implement the narrow band system the audio passband filters which select the formants would be required to have skirts much steeper than the filters used in the base

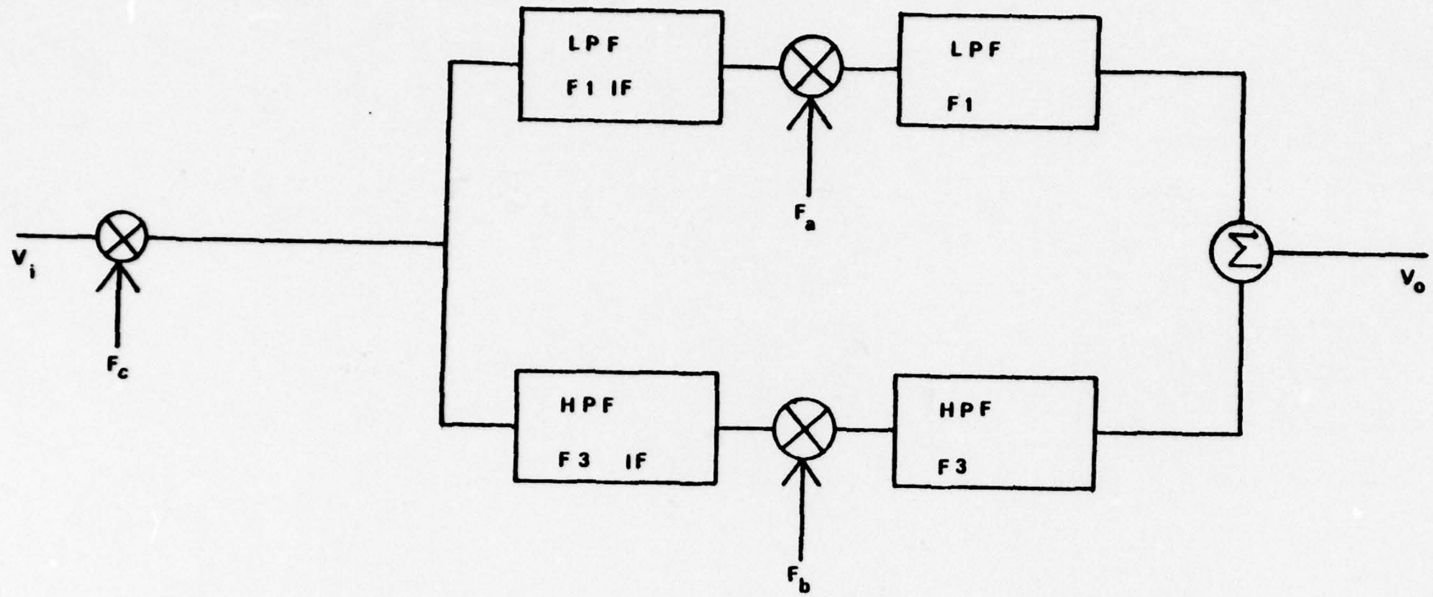


FIGURE 19 RECEIVER PROCESSOR

band test. Ideally the compressed spectrum would have a cross-over point which was at least -15dB down. This would result in minimal spurious mixing due to overlap.

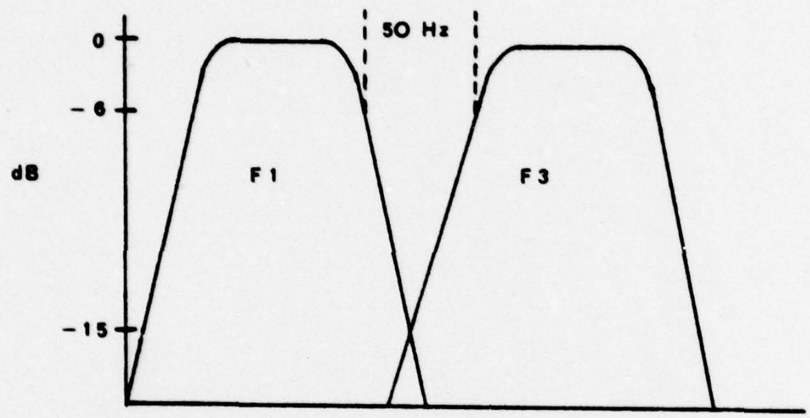


FIGURE 20 COMPRESSED SPECTRUM

It was determined that 27th order filters would meet this requirement. The filters can easily be constructed using hybrid monolithic integrated circuits with laser trimmed resistors. By use of this technique cutoff frequencies can be precisely set and most of the tolerance and mismatch problems associated with a discrete component realization can be eliminated. Feedback and groundloop are all but eliminated.

Using this same technique for construction of the 750Hz low pass filter and the 1950Hz high pass in the receiver adapter insures that the unwanted spectral components of the adjacent formants which may have passed through the IF filters will be suppressed. This is shown in Figure 21.

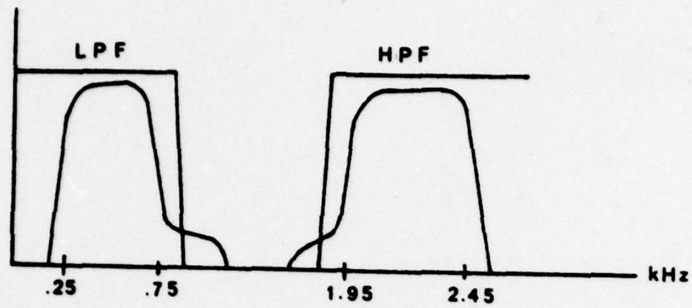


FIGURE 21

SEPARATION FILTERS

A block diagram of the complete system is shown in Figure 22.

C. APPLICATIONS OF THE NARROW BAND VOICE SYSTEM

Applications of the narrow band voice system were developed for both analog voice and digital voice transmission. One principal characteristic of the narrow band voice system is that it is external to existing equipment. No equipment modifications are necessary to employ narrow band voice modulation. It is only necessary to provide a processor at each end of the channel.

The principal advantage of narrow band voice modulation is that more users can communicate simultaneously in a given channel. Consider a standard analog communication channel of band width 200 to 2500Hz. By proper selection of the IF frequencies in the voice processors, two compressed voice signals can be transmitted in this channel. This is shown in Figure 23.

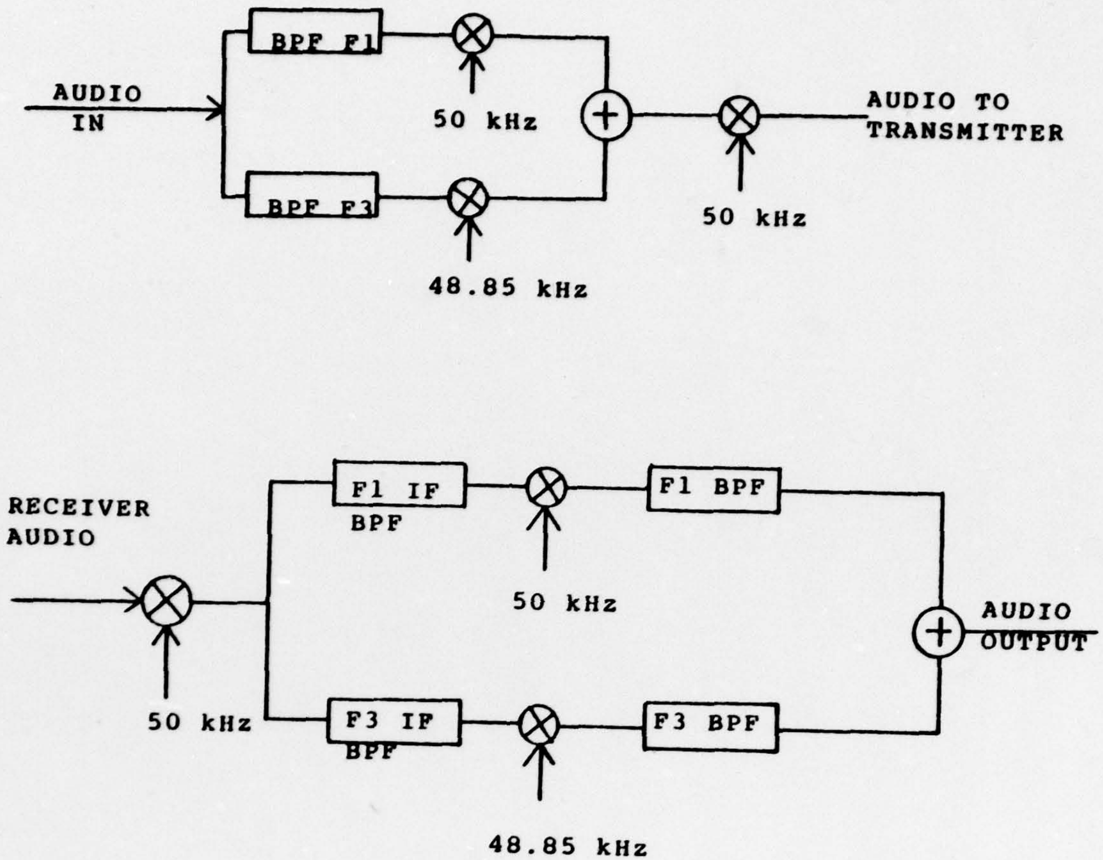


FIGURE 22 NARROWBAND VOICE SYSTEM

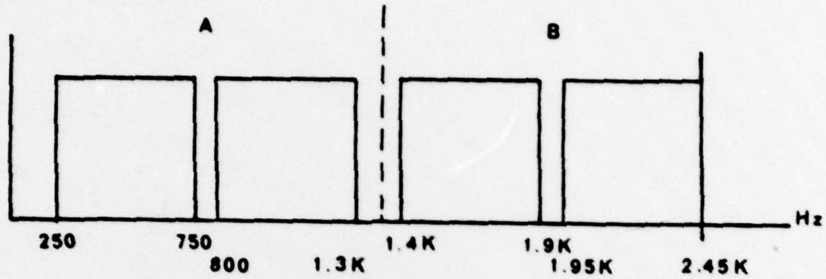


FIGURE 23 APPLICATION OF NARROWBAND VOICE

Considering the radio frequency transmission of voice, such as HF, yields similar results in bandwidth reduction. With this system the transmission bandwidth of the audio signal is approximately the carrier frequency ±1300Hz for DSB/AM. For single sideband, the bandwidth required is 1300Hz. Comparing these figures to 5000Hz for DSB/AM and 2500Hz for SSB in a standard system shows an approximate reduction in bandwidth of 50 percent. Hence twice as many users can occupy an RF spectrum using narrow band voice modulations.

Digital communications are also enhanced by using the narrow band technique. For a signal which is band limited to 2500Hz and sampled at a typical rate slightly above the Nyguist rate, the frequency of the sampler would be 6000Hz. If analog to digital conversion results in an 8 bit word for each sample amplitude, as is done in telephone work, 48,000 bits per second are required. Considering the narrow band voice

signal which is band limited to 1300Hz, and sampling in the same manner, a sampling frequency of 3000Hz would be used. This results in 24,000 bits per second being required, a reduction of 50 percent. Clearly twice as many users could be accommodated by existing equipment.

D. CONCLUSIONS AND RECOMMENDATIONS

It has been shown in this thesis that the bandwidth required for highly intelligible voice signals is approximately 1300Hz. Formants F1 and F3 can be isolated and compressed into a narrow band voice signal. A system was developed in which outboard devices process voice at the transmitter and receiver to allow reduction of the required transmission spectrum.

There are two areas of research which are recommended for further investigation in conjunction with this thesis. First, the degree to which hybrid monolithic active filters can be used to isolate the formants is still unclear. There is a possibility that orders of filter as high as 40 may be realized using this technique. If so, the overall spectrum may be reduced even further by elimination of the 50Hz guard band between the formants and the elimination of the wasted spectrum from 0 to 250Hz.

Secondly, the magnitude of the distortion caused by overlap of the formants in the compressed spectrum has not been fully investigated. It is hypothesized that by maintaining the crossover point 15dB down, the distortion will be minimal, but this has not been proved. An additional area of interest would be the intelligibility of the compressed baseband signal

when the intercept receiver is not equipped with a narrowband voice processor. It is hypothesized that the signal would be almost unintelligible since the normal harmonic relationships inherent in normal voice signal would be destroyed.

The problem of accommodating the increasing number of subscribers given the constraints of bandwidth and costs associated with new equipment can be solved by the narrow band voice system developed in this thesis. This technique offers a noble alternative to the more costly and complex cellular land radio systems presently under consideration by the FCC.

APPENDIX A:

COMPUTER PROGRAMS IN SUPPORT OF THE DATA ANALYSIS

Data recorded on tape by the digital-analog hybrid computer was in a 7-track, 556 BPI format. In order to use this data on the IBM 360/67, the tape had to be converted to a 9-track format. This was accomplished using the program listed below.

```
      DIMENSION IDAT(4096),DAT(4096)
      FACTOR=100./(2.**23)
      REWIND 2
      REWIND 4
      M=0
      LRECL=4096
11    J=0
      READ(2,15,END=50)
      READ(2,15,END=50)
10    READ(2,15,END=50,ERR=60)IDAT
15    FORMAT(128(8A4))
      J=J+1
      WRITE(6,70)J
70    FORMAT('0',10X,'RECORD NO.=',I4)
      CALL FORM(IDAT,LRECL)
      DO 22 I=1,LRECL
22    DAT(I)=IDAT(I)*FACTOR
      WRITE(4,15)DAT
      IF(J.LT.3) GO TO 10
      GO TO 50
60    WRITE(6,61)J
61    FORMAT('0',5X,'READ ERROR, RECORD NUMBER=',I3)
      GO TO 10
50    WRITE(6,51)M,J
51    FORMAT('0',5X,'END OF FILE',I2,'NO. RECORDS=',I3)
      END FILE 4
      M=M+1
      IF(M.LT.7) GO TO 11
      WRITE(6,71) M
71    FORMAT('0',5X,'END OF TAPE, FILES=',I3)
      RETURN
      END
```


Computation of the DFT components was accomplished by use of HARM, a prepared subroutine from the IMSL Library. The main program written to compute, list, and plot the DFT components is as shown below.

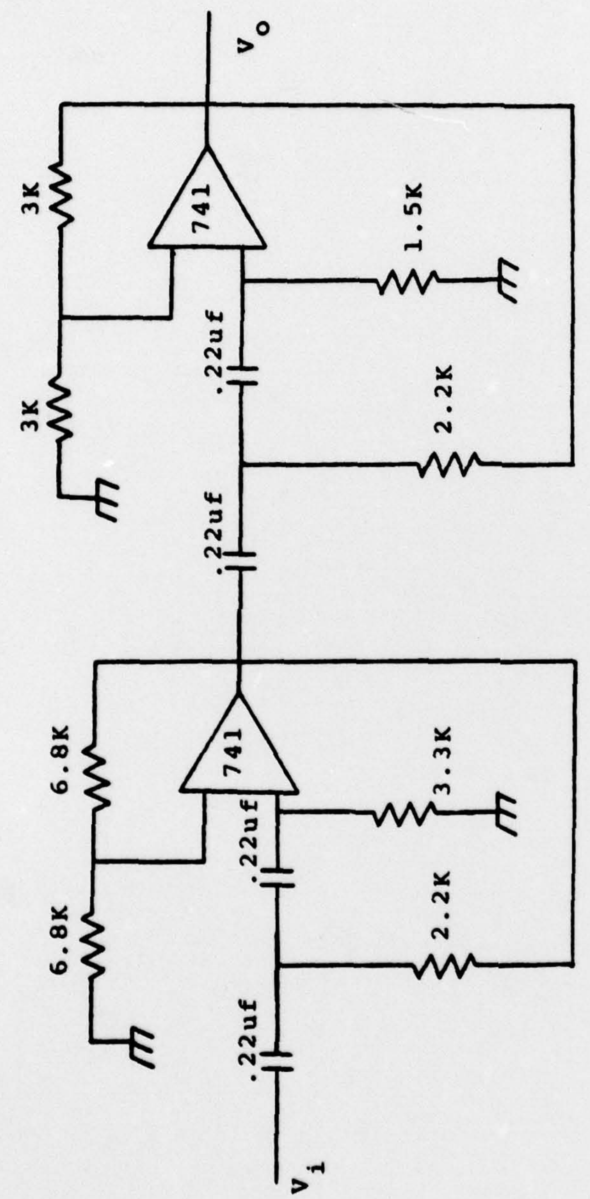
```
      DIMENSION S(1024),INV(1024),M(3)
      DIMENSION WH(1024),W(1024)
      DIMENSION G(1024),T(600)
      DIMENSION RANGE(4),X(80),Y(80)
      COMPLEX*8 A(1024,1,1),WOFT(1024,1,1)
C
C
      NR=7
      KK=10
      N1=2**KK
      N2=1
      N3=1
      NH=N1/2
      M(1)=KK
      M(2)=0
      M(3)=0
C
C
C
      IR=1
      DO 197 KJ=1,512
197   T(KJ)=0.0
      REWIND 2
      DO 178 NR=1,3
C     READ IN THE DATA
      READ(2,15)WH
15    FORMAT(128(8A4))
C
C
C
C     TRANSFER THE WH(K) DATA TO THE COMPLEX A MATRIX
C
C
      DO 3 I1=1,N1
    3 A(I1,1,1)=WH(I1)
      CONTINUE
C
C
C     COMPUTE THE DISCRETE FOURIER TRANSFORM OF THE
C     WH(K) DATA
C
C
      CALL HARM(A,M,INV,S,1,IFERR)
C
C
C     CALCULATE THE MAGNITUDE OF THE FOURIER COMPONENTS
C     AND STORE ONLY THE POS FREQ VALUES OF W(I1)
C
C
      DO 30 I1=1,NH
   30 W(I1)=CABS(A(I1,1,1))
C
C
C     PRINT THE DFT COEFFICIENTS
C
      WRITE(6,211)NR
211   FORMAT(1X,'FILE NUMBER',I3)
      DO 61 I1=1,NH
61    W(I1)=W(I1)/129.
      DO 190 NM=1,512
190   T(NM)=T(NM)+W(NM)
      L=129
      J=257
      JJ=385
      DO 62 N=1,128
      WRITE(6,101)N,W(N),L,W(L),J,W(J),JJ,W(JJ)
      L=L+1
      J=J+1
   62 JJ=JJ+1
  101 FORMAT(I4,4X,F10.6,4X,I4,4X,F10.6,4X,I4,4X,F10.6,4X,I4,4X,F10.6)
C
C
C     PLOT THE SPECTRUMS
C
```

```
C
      RANGE(1)=3200.0                                              DFT00800
      RANGE(2)=200.0
      RANGE(3)=3.0
      RANGE(4)=0.0
      N=40                                                         DFT00870
      K=1                                                          DFT00880
      MODCUR=0                                                     DFT00890
      AN=200.0
      DO 160 I=1,40
      X(I)=AN
160   AN=AN+74.0
      KN=27
      DO 170 I=1,80
      G(I)=W(KN)
170   KN=KN+5
      NN=1
      DO 175 L=1,40
      Y(L)=G(NN)
175   NN=NN+2
      DIS=1.0
      SC=0.0
      MAXIT=40
      WK=56.
      NX=N
      CALL ICSMOU(X,Y,NX,DIS,SC,MAXIT,WK,IER)
      WRITE(6,161)                                                 DFT00950
  161 FORMAT(1H1)                                                  DFT00960
      CALL UTPLOT (X,Y,N,RANGE,K,MODCUP)                           DFT00980
      WRITE(6,162)IR
162   FORMAT(1X,'GRAPH OF SAMPLES FOR RECORD NUMBER',I5)
      WRITE(6,161)
178   IR=IR+1
      WRITE(7,211)NR
      DO 191 N=1,512
191   T(N)=T(N)/3.0
      L=129
      J=257
      JJ=385
      DO 195 N=1,128
      WRITE(6,101)N,T(N),L,T(L),J,T(J),JJ,T(JJ)
      WRITE(7,301)N,T(N),T(L),T(J),T(JJ)
301   FORMAT(I4,4F10.6)
      L=L+1
      J=J+1
195   JJ=JJ+1
```

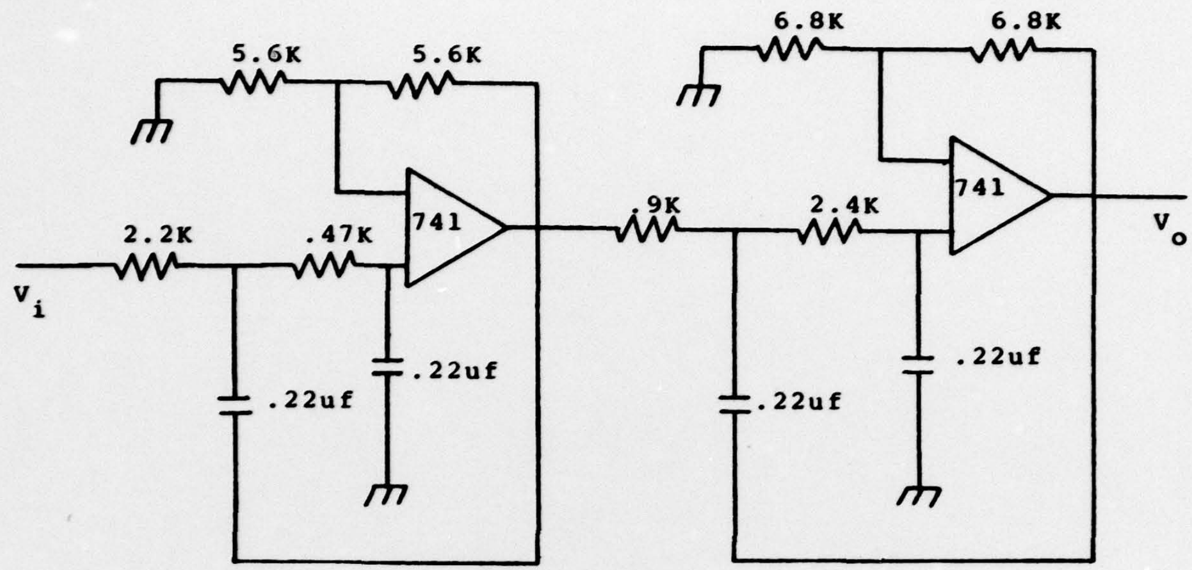
Computation of the power spectral density and plotting was accomplished by the program listed below.

```
      DIMENSION G(512),T(512),X(512),Y(512),RANGE(4),Z(512)
      ID=1
      L=129
      J=257
      JJ=385
1     DO 10 N=1,128
      READ(5,401)N,T(N),T(L),T(J),T(JJ)
401   FORMAT(I4,4F10.6)
      L=L+1
      J=J+1
10    JJ=JJ+1
      L=129
      J=257
      JJ=385
      KN=27
      AN=200.0
      DO 198 I=1,40
      X(I)=AN
198   AN=AN+74.0
      DO 201 I=1,80
      G(I)=T(KN)
201   KN=KN+5
      NN=1
      DO 202 L=1,40
      Y(L)=G(NN)
202   NN=NN+2
      RANGE(1)=3200.0
      RANGE(2)=200.0
      RANGE(3)=3.0
      RANGE(4)=0.0
      N=40
      K=1
      MODCUR=0
      DIS=1.0
      SC=0.0
      MAXIT=40
      WK=56.
      NX=N
      CALL ICSMOU(X,Y,NX,DIS,SC,MAXIT,WK,IER)
      WRITE(6,161)
      CALL UTPLOT(X,Y,N,RANGE,K,MODCUR)
      WRITE(6,193)
193   FORMAT(1X,'GRAPH OF THE AVERAGES')
      WRITE(6,161)
      DO 12 LL=1,512
12    Z(LL)=20.*(ALOG10(T(LL)))
      DO 17 N=1,128
      WRITE(6,101)N,Z(N),L,Z(L),J,Z(J),JJ,Z(JJ)
  101 FORMAT(I4,4X,F10.6,4X,I4,4X,F10.6,4X,I4,4X,F10.6,4X,I4,4X,F10.6)
      L=L+1
      J=J+1
17    JJ=JJ+1
      AN=200.
      DO 14 N=1,80
      X(N)=AN
14    AN=AN+37.5
      KN=27
      DO 15 I=1,80
      Y(I)=Z(KN)
15    KN=KN+5
      DIS=1.0
      SC=0.0
      MAXIT=80
      WK=136.0
      NX=80
      CALL ICSMOU(X,Y,NX,DIS,SC,MAXIT,WK,IER)
      RANGE(1)=3200.0
      RANGE(2)=200.0
      RANGE(3)=10.0
      RANGE(4)=-30.0
      N=80
      K=1
      MODCUR=0
      WRITE(6,161)
161   FORMAT(1H1)
      WRITE(6,162)
162   FORMAT(1X,'GRAPH OF THE FINAL AVERAGES   LOG SCALE')
      WRITE(6,165)ID
165   FORMAT(1X,'FILE NUMBER',I4)
      CALL UTPLOT(X,Y,N,RANGE,K,MODCUR)
      WRITE(6,161)
      ID=ID+1
      IF(ID.NE.7) GO TO 1
      STOP
      END
```

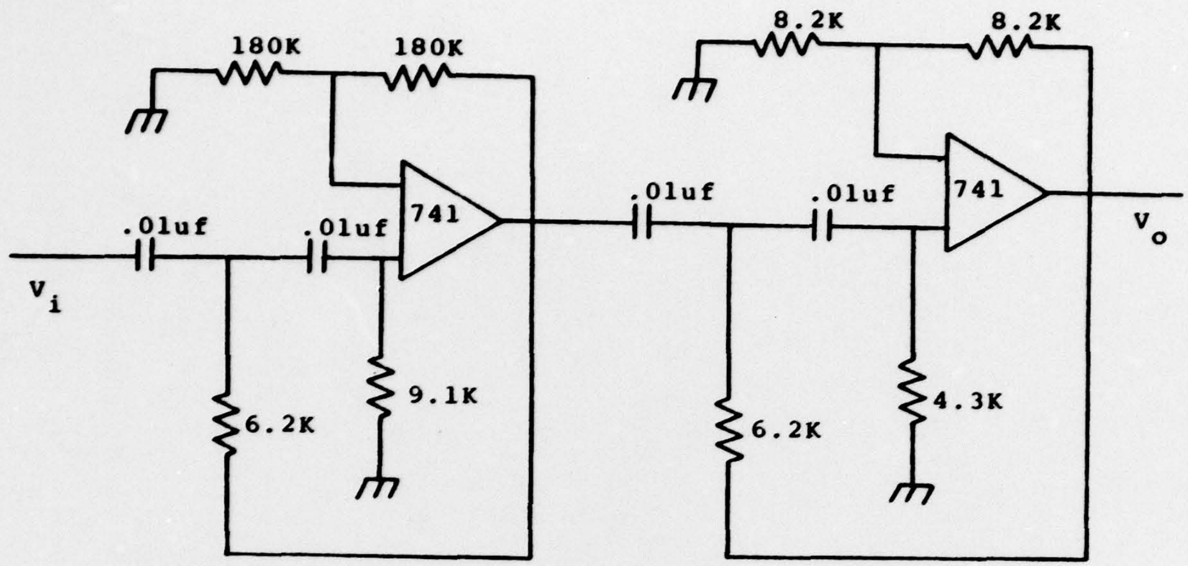
<u>APPENDIX B: SCHEMATIC OF THE FILTERS</u>



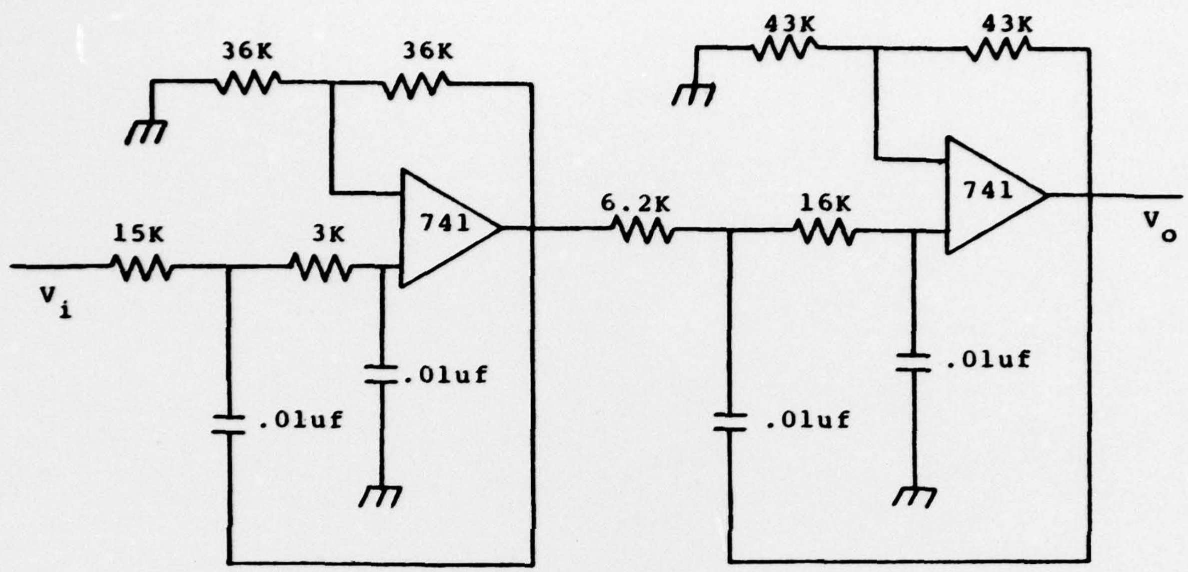
250 Hz HIGHPASS FILTER



650 HZ LOWPASS FILTER



1950 Hz HIGHPASS FILTER



2450 Hz LOWPASS FILTER

LIST OF REFERENCES

1. Weber, M., "An Effective Bandwidth Reduction for Communications Use in a VHF/UHG Satellite Link," IEEE Trans. on Audio and Electroacoustics, v. AU-17, pp. 222-223, September 1969.

2. Campanella, S. J., "A Survey of Speech Bandwidth Compression Techniques," IRE Trans. on Audio, v. 5, pp. 104-106, September-October 1958.

3. McCormick, E.J., "Human Factors in Engineering and Design," pp. 142-156, McGraw Hill, 1976.

4. Fletcher, H., Speech and Hearing in Communication, pp. 135-136, Van Nostrand, 1953.

5. Kryter, K. D., "Speech Bandwidth Compression Through Spectrum Selection," J. Acoust. Soc. Am., v. 32, pp. 547-556, May 1960.

6. Harris, R. W. and Cleveland, J.F., "A Baseband Communications System," QST, v. LXII, pp. 14-21, December 1978.

7. Olson, H. F., "Speech Processing Techniques and Applications," IEEE Trans on Audio and Electroacoustics, v. AU-15, pp. 120-126, September 1967.

8. Daguet, J. L., "Speech Compression CODIMEX System," IEEE Trans. on Audio, pp. 63-71, March-April 1963.

9. Richards, D. L., "Statistical Properties of Speech Signals," Proc. IEEE, v. III, pp. 941-947, May 1964.

10. Lancaster, D., Active Filter Cookbook, Howard W. Sams, 1975.

11. Dudley, H., "The Carrier Nature of Speech," Bell System Tech. J., v. 19, pp. 495-515, 1940.

12. Fant, G., The Acoustics of Speech, paper presented at Third International Congress on Acoustics, 1959.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93940	2
3.	Department Chairman, Code 62 Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	1
4.	Associate Professor G. D. Ewing, Code 62 Dk Department of Electrical Engineering Naval Postgraduate School Monterey, California 93940	1
5.	CPT Raymond R. Hitchcock, USA 3145 Heatherridge Lane Reno, Nevada 89509	1